DOCUMENT RESUME

ED 193 554 CG 014 715

AUTHOR Horan, John J.
TITLE Experimentation in Counseling and Psychotherapy: New and Renewed Mythologies.
PUB DATE Apr 80
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Counseling: *Counseling Effectiveness; *Evaluation Criteria; *Literature Reviews; *Psychological Studies; Psychotherapy; *Research Methodology; *Research Problems; State of the Art Reviews; Statistical Analysis

ABSTRACT

Five myths permeate the counseling and psychotherapy literature. Two of these myths concern the absolute and differential effects of counseling and psychotherapy. Contrary to the conclusions of Smith and Glass's (1977) meta-analysis, it has not been established that aggregated counseling and psychotherapy schools are effective at all, much less are they equally effective. The remaining three myths are more fundamental: our experimental subjects often do not receive treatments appropriate to their clinical problems; our treatments are frequently not deployed as purported; and finally our control groups rarely address one of the most powerful artifacts of all. (Author).

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED193554

Experimentation in Counseling and Psychotherapy: New and Renewed Mythologies

John J. Horan

The Pennsylvania State University

Five-Presidential State-of-the-Art Address to the American Educational Research Association, Boston, April, 1980.

Requests for reprints should be sent to the author at 315 Carpenter Building, University Park, Pennsylvania 16802.

CG 014715

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John J. Horan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Experimentation in Counseling and Psychotherapy:

New and Renewed Mythologies

In his 1976 presidential address to the American Educational Research Association, Gene Glass suggested that we have found ourselves in "the mildly embarrassing position of knowing less than we have proven." He coined the term "meta-analysis" to refer to a particular method of extracting information from large accumulations of individual studies. In this talk and in subsequent publications, Glass and his colleague Mary Lee Smith (Glass, 1978; Smith & Glass, 1977) applied meta-analysis to a large population of counseling and psychotherapy outcome studies.

As you might well imagine, I spent a good deal of time over the past year obsessing on the contents of my talk today. At the outset I had no intention whatsoever of dealing with the psychotherapy meta-analysis, but like a newly-blossomed nose blemish I found it hard to ignore. Since no legitimate state-of-the-art address could avoid dealing with this particular meta-analysis, I decided to bite the bullet, confront it head on, and use its Achilles' heel to introduce the topics of greatest concern to me.

In contrast to Glass it is my belief that we find ourselves in the terribly embarrassing position of having proven far less than we purport to know. There is a quantum leap between our experimental literature and our methodological sophistication. We now know what's wrong with our data, but too many of us pretend to our students and to our public that there is solid empirical evidence behind our varied proclamations. Like the seers of ancient Greece we perpetuate our own Olympian myths with the most specious of arguments rather than admit our ignorance of the natural phenomena in question.

In the space of an hour I cannot hope to recite the entire anthology of fairy tales that pervade our profession. Instead I will focus on five myths that I believe are classic examples of self-deceit. Two of these myths concern the absolute and differential effects of psychotherapy. Many of us have been deluded into thinking that our literature has clearly established the facts that aggregated psychotherapies do indeed work and that the various individual psychotherapies are equally effective. Both of these myths were actually given a fiery demise long ago by Kiesler (1966) and Krumboltz (1968); but like the fabled phoenix they now rise from their own ashes in the colorful feathers of meta-analysis.

The remaining three myths are of more recent vintage. We believe and profess that the subjects in our experiments receive treatments appropriate to their clinical problems, that the treatments are in fact deployed as purported, and that our customary control groups allow us to determine the existence of a treatment effect. After reviewing the renewed myths of meta-analysis, I will discuss each of these new myths in turn.

The Renewed Myths of Meta-analysis

Background

For research endeavors in the history of counseling have met with as much vitriolic, glowing, and satiric commentary as has meta-analysis. Eysenck (1978), for example, referred to it as "an exercise in mega-silliness." Scriven (1979), in contrast, elevated Glass's work to the status of "the definitive study in the field." Finally, with tongue in cheek, three nom de plums (Kazrin, Durac, & Agteros, 1979, p. 398) proposed an improved methodology called "meta-meta analysis" which spares the consumer the "onerous effort of reading individual studies" because all the necessary information "can be obtained from a journal's table of contents."

The basic unit of meta-analysis is a score called "effect size" defined

as the mean difference between the treated and control subjects devided by the standard deviation of the control group. Thus "effect size" is essentially a z score representing the degree of success produced by an experimental treatment on a specific measure in a given experiment at a particular point in time. In the Smith and Glass (1977) meta-analysis of the psychotherapy outcome literature, 375 studies which produced 833 effect-size scores were further classified in terms of 16 contextual variables including type of therapy, therapist experience, subject pathology, and the internal validity of the research design. From this amorphous mass of data they attempted to show that a) psychotherapies, on the whole, are beneficial, and b) the various individual psychotherapies are equally beneficial. In so doing they perpetuate "The Absolute Effectiveness Myth" and "The Comparative Equality Myth."

The Absolute Effectiveness Myth

Smith and Glass (1977) offer essentially one argument in support of The Absolute Effectiveness Myth: Psychotherapies on the average produce .68 of a standard deviation of improvement on all measures relative to control subjects, or in other words, comparative movement from the 50th to the 75th percentile. A less optimistic way of viewing this change was offered by Gallo (1978) who analyzed the Smith and Glass data in a different manner and concluded that only "10% of the total variance in the adjustment scores of both treatment and control patients could be accounted for by the effects of psychotherapy" (p. 515). Certain aspects of the original meta-analysis data are equally damaging to the claim that aggregated psychotherapies are beneficial. Effect size, for example, correlated -.02 with duration of therapy, and -.01 with experience of therapist. Thus, as Rimland (1979, p. 192) points out "a client can expect just as much benefit from consulting an untrained lay person for one session as he or she can from consulting a highly trained M.D. or Ph.D. for many hundreds of (expensive) hours."

To the sobering observations of Gallo and Rimland I would like to add another note of gloom. The psychotherapy meta-analysis did not differentiate control conditions involving no treatment at all from control conditions invoking high demand characteristics. Indeed the studies that rated highest on the three-point "quality-of-design" scale met only two criteria: randomization and low mortality. In effect, Smith and Glass "graded on the curve," giving "As" for excellence to many projects that were at best mediocre! We all know it is fairly easy to show that one's favorite therapy is better than nothing, but quite difficult to demonstrate its superiority to a highly credible, but theoretically-inert alternative treatment. More about this later! For now I think it's safe to assume that most studies in this prior-to-1976-population paid scant attention to the "quality" of the control treatment (see, for example, Kazdin & Wilcoxon, 1976) Hence, the average improvement score of the psychotherapy meta-analysis (.68 of a standard deviation) probably represents little more than the placebo phenomenon in various guises. Mind you, I'm not echoing Scriven's (1979) bald assertion that psychotherapy is nothing more than a process of raising client hopes; I'm simply saying that the population of psychotherapy outcome studies up to 1976 hardly offers convincing evidence to the contary.

The Comparative Equality Myth

Much of the controversy generated by the psychotherapy meta-analysis, however, concerns its perpetuation of The Comparative Equality Myth. Glass (1976, p. 7) argues thusly: "For all the superiority claimed by one camp or the other, for all the attention lovingly squandered on this style of therapy versus that style, the available evidence shows essentially no difference in the average impact of each class of therapy." Smith and Glass (1977) offer three different analyses in support of this myth.

First, the ten different adjectives originally used to describe the

various psychotherapies were collapsed into two therapy "super-classes" called behavioral and nonbehavioral. The behavioral therapies reportedly yielded an average effect size of .8 of a standard deviation in contrast to the nonbehavioral average effect of .6. Because the behavioral therapies allegedly used more "subjective" outcome measures and shorter follow-up periods, Smith and Glass (1977, p. 758) suggest that the .2 of a standard deviation difference "is somewhat exaggerated in favor of the behavioral superclass."

The second analysis involved only those studies ($N \approx 50$) in which a behavioral therapy was simultaneously compared to a nonbehavioral therapy. Here, the superiority of the behavioral superclass reportedly shrinks to only .07 of a standard deviation. Smith and Glass feel that this analysis is particularly convincing since the context of each study was equivalent for the two superclasses.

The third route to the relative equality conclusion consisted of several regression analyses. Smith and Glass initially observed that their contextual variables produced a multiple correlation of .50 with effect size. By setting these predictor variables to specified points (for example, highly intelligent phobic clients seen by therapists with two year's experience), they believed they could estimate the effect produced by a particular class of therapy in a given set of circumstances. In two prototypical examples offered, the behavior therapies reportedly showed a trivial superiority over the psychodynamic approach.

In tracing the flow of these three arguments supporting The Comparative Equality Myth, I am reminded of a rather paranoid individual I once knew--his logic was absolutely impeccable, but his assumptions were simply incredible!

I would like to call your attention to several problematic assumptions underlying the logic of this particular meta-analysis.

1. The vegetable soup assumption. Smith and Glass (1977, p. 753) insist that "mixing different outcomes together is defensible" primarily because "all outcomes are more or less related to 'well being' and so at a general level are comparable." In other words the shrinking wet spots of enuretic kids have the same meaning as, and can be averaged with, the gains in happiness self-reported by outpatient college students. Moreover, these variables are not only equally related to each other but also to everything else from assertiveness to self-concept to orgasmic frequency. The researcher who is not troubled by this assumption is certainly encouraged to avoid recalcitrant clinical problems and exploratory generalization criteria, otherwise the average effect size score will shrink enormously and contribute to a poor showing for his or her side in the next meta-analysis.

2. The therapy-uniformity assumption. The psychotherapy meta-analysis erroneously assumes that the behavioral and nonbehavioral therapies can be treated as superclasses and meaningfully compared with each other. Such wholesale reductionism buries extremely important differences that vividly emerge under a more finely grained analysis (Kazdin & Wilson, 1978). For example, within the so-called behavioral superclass performance-based strategies such as participant modeling are clearly superior to the traditional imagery-based version of systematic desensitization (Bandura, 1976; Bandura, Blanchard, & Ritter, 1969). Moreover, parametric evaluations of flooding and stress inoculation demonstrate that simple procedural variations may enhance the outcome of a given technique (Stern & Marks, 1973; Sherry & Levine, 1980; Horan, Hackett, Buchanan, Stone, & Demchick-Stone, 1977; Hackett, Horan, Buchanan, & Zumoff, 1979; Hackett & Horan, 1980). In the comparison of superclasses, the psychotherapy meta-

analysis not only ignores important within-class differences but also gives equal weight to the failing techniques and experimental wastelands of bygone eras.

Even if we accept the therapy-uniformity assumption, the ingredients of the meta-analysis superclasses warrant closer inspection. Rational-emotive therapy was included in the nonbehavior grouping; gestalt therapy (which they judged incredibly as similar to the behavioral perspective!) was excluded from the analysis altogether. Moreover, studies on implosion (a technique muled from the promiscuous union of psychoanalysis and the animal laboratory) were stuffed into the behavioral package. In point of fact, gestalt therapy should have been included with the nonbehavioral techniques, and rational-emotive therapy should have been placed with the behavioral procedures (Ellis, 1977; Horan, 1979; Mahoney, 1974; Meichenbaum, 1977; Presby, 1978). In fairness to both superclasses, studies dealing with the thoroughly discredited implosion technique should have been excluded from the meta-analysis (see Morganstern, 1973). After making these minor adjustments I reanalyzed the Smith and Glass (1977) data relative to the first argument supporting the Comparative Equality Myth and found an appreciable increase in the superiority of the behavioral superclass.

3. The bad-data-is-valuable assumption. Smith and Glass have been frequently and unfairly roasted for advocating the collection of bad data. To the contrary, their position has always been that one's next study ought to have the best possible design, but once the study is completed it becomes an empirical question whether poorly designed studies yield results at odds with those of well designed studies. Should such a finding occur, then design quality can be used as a covariate when comparing the effects of two treatment classes. In the psychotherapy meta-analysis there was a significant

relationship between design quality and effect size ($p < .05$); however, Glass chose (1978, p. 3) to judge it small enough to warrant the following conclusion: "For this large body of research, it is an empirical fact that 'good' and 'bad' studies show the same results."

The controversy here essentially reduces to three questions: a) Did the meta-analysis demonstrate the fact of no relationship? b) Can such a relationship be meaningfully represented in terms of a simple correlation coefficient? c) Is it legitimate to use analysis of covariance in this manner? As we shall now see, the answer to all three questions is "no."

a) Glass is correct in asserting that the relationship between effect size and design quality is an empirical question, but by no stretch of the imagination has he demonstrated the empirical fact of no relationship. I again call your attention to the three-point scale for evaluating design quality. The items were: (1) High--randomization and low mortality, (2) Medium--more than one threat to internal validity, and (3) Low--no matching of pretest information to equate groups.

At best such a scale can only discriminate marginally adequate studies from totally inadequate ones. It does not speak meaningfully to the matter of design quality. Certainly one can ask whether wormy apples are more palatable than completely rotten ones, but the real food for thought concerns which relatively blemish-free variety makes the better pie. Had the meta-analysis included only those studies having some claim to internal validity (the sine qua non for feeling confidence in one's data), and had the quality-of-design scale been constructed in such a way as to discriminate the really good studies from the barely adequate ones, perhaps the meta-analysis would have yielded a different conclusion.

b) It is indeed doubtful that the relationship between design quality and effect size can be adequately expressed in terms of a simple correlation

coefficient. To illustrate this problem let us consider some possible outcomes of both poorly designed and well designed studies. In the case of poorly designed studies effect size would probably be related to the nature of the design flaws, not their frequency. For example, failure to adequately control placebo influences may yield a spuriously high effect size; failure to employ reliable measures may result in an artifically low effect size. Summing these "offsetting" flaws would obscure the real relationship. In the case of well designed studies effect size is presumably related to the actual lawfulness of the phenomena being investigated. Thus, good studies of powerful treatments should yield large effect size scores; equally good studies of weak treatments should produce low effect size scores. Reality is independent of one's choice of experimental design!

Essentially then, effect size is determined by specific kinds of design flaws (not their total) and also by conditions of nature. It seems unlikely that a stable overall correlation between design quality and effect size will ever be found; and if indeed one did emerge, its meaning would be lost against the backdrop of complex underlying processes.

c) The final flaw in the bad-data-is-valuable assumption concerns the legitimacy of using design quality as a potential covariate when comparing the effects of two or more psychotherapy treatment classes. Analysis of covariance (ANCOVA) is a very valuable tool for improving the power of experimental designs where subjects are randomly assigned to treatments. Meta-analysis, however, constitutes a quasi-experiment in which the treatment classes are essentially organismic variables (i.e., not randomly assigned). The appropriateness of ANCOVA in such situations has been questioned by Games (1976) and others (Cronbach & Furby, 1970; Lord, 1969).

In ail of this, I sincerely hope my remarks are not construed as a personal assault on Gene Glass's competence. He is a well-respected scholar and one of the foremost statisticians in our profession. Indeed, it is primarily because of his stature that meta-analysis has received such a high degree of attention. My own comments are directed toward this particular meta-analysis of the psychotherapy literature, not to the future of the technology nor to its application in other areas. I would like to move on now to three additional myths which inhibit our understanding of the counseling and psychotherapy outcome literature.

New Myths About Old Realities

The Appropriate Treatment Myth

Publication of Campbell and Stanley's (1966) classic little book on experimental design had an enormous impact on the field of counseling and psychotherapy. Even today, dissertation proposals which don't quite fit the experimental mold are viewed with a jaundiced eye. It is regrettable that we don't have a similar Campbell and Stanley "bible" to guide our empirical conduct in the area of clinical problem definition. Most counseling outcome studies rest on subject screening criteria or pretest measures that may give the illusion of rigor but in fact provide insufficient information on which to make an appropriate treatment decision. Consequently, a substantial percentage of any subject pool invariably receives a counseling intervention that is theoretically irrelevant to their actual clinical problem. Let me cite several examples.

Many subjects are operationally labeled "phobic" because they refuse to approach or handle a snake, spider, rat, or whatever, and their verbal reports

indicate a similar reluctance. Although such avoidance behavior is typical of truly phobic subjects, it is also displayed by subjects who are adaptively skeptical. (BAT test animals may be nonpoisonous, but they are perfectly capable of biting.) Furthermore, this particular operational umbrella covers a goodly number of nonphobic people who essentially have erroneous (and not so erroneous) beliefs about the animal such as its being slimy, dirty, or a carrier of disease. Still other subjects will have widely differing degrees of fear, skepticism, and mistaken belief in combination. A treatment such as desensitization would be theoretically appropriate only to the truly phobic characteristics of a subject pool, and these characteristics may be minor or possibly even nonexistent.

In the counseling and psychotherapy literature the only thing more common than small animal phobia studies are complaints about such studies (see, for example, Barrios, 1977; Cooper, Furst, & Bridger, 1969; Bernstein & Paul, 1971). Perhaps the Appropriate Treatment Myth would be better illustrated with the clinical problem of test anxiety. We are all aware of the classic 1908 Yerkes-Dodson law that posits a curvilinear relationship between anxiety and performance. This law suggests that a moderate amount of test anxiety may be quite helpful to students desiring higher grades. Strictly speaking then, we cannot assume that a clinical problem of maladaptive test anxiety exists, unless we can document that the anxiety produced by "testing stimuli" in turn yields lowered performance levels. I have yet to encounter a counseling outcome study which clearly established the existence of such maladaptive test anxiety in its subject pool.

Be that as it may. Even if we assume verbal reports or the act of volunteering for treatment to be sufficient grounds for the establishment of

maladaptive test anxiety[1]; there are still many different clinical problems that fall under this generic label and no single treatment is appropriate to all of them. Cue-controlled relaxation, for example, might be theoretically relevant to the acute anxiety experienced by a previously unanxious high achiever who now faces an entrance examination for a professional school. It would probably be inappropriate, however, for students with deficient reading or study skills whose self-reported test anxiety is a consequence rather than a cause of chronic poor performance. Moreover, any treatment other than cognitive restructuring would have highly debatable relevance to the anxious perfectionist who believes that a less than "curve-setting" performance would be absolutely catastrophic. Finally, at least one form of test anxiety is essentially untreatable, namely the natural consequence of a decision to play instead of to study.

In the counseling and psychotherapy literature, other examples of inappropriate treatments applied to crudely defined clinical problems abound. Many instances of "unassertiveness," for example, are really decision making concerns rather than skill deficits (see Fiedler & Beach, 1978). Thus, the frequently deployed procedure called "behavioral rehearsal" would be irrelevant to subjects who can already act in an assertive, or extinguishing, or polite, or empathic manner, but who adaptively wonder which response pattern will maximize the probability of getting promoted, making a sale, salvaging a familial relationship, or acquiring some other utility. To paraphrase the words of my good friend and colleague George Hudson, even in university settings supposedly characterized by higher levels of rationality and receptivity to honest communication, it sometimes shows a fine command of the language to say nothing!

The Appropriate Treatment Myth owes its existence to two common lapses of thought. The first involves the erroneous assumption that because we have a baptismal name for our screening criteria or dependent measures, we therefore must be assessing a homogeneous clinical concern. From this precarious cognitive precipice it is but a short hop to the equally mistaken belief that because our favorite counseling intervention may be theoretically linked to a particular form of that problem, it must consequently be relevant to the entire subject pool. In point of fact virtually all clinical problems mentioned in the titles of our journal articles are essentially crude general descriptions of specific client concerns that probably require differential treatment.

Failure to recognize the Appropriate Treatment Myth has three serious consequences. In the first place, inclusion of subjects whose actual clinical problem is irrelevant to the experimental treatment will lower or indeed wash out the average impact of that treatment. Even if the study is fortunate enough to escape this particular type of error, the emerging "significant" gain will inevitably be trivial by clinical standards. Though science does indeed advance by small steps, our counseling and psychotherapy literature is plagued by artifacts that are as frequent and as powerful as our most effective treatments (see Barber, 1976; Badia, Haber, & Runyon, 1970; Rosenthal & Rosnow, 1969). I for one would derive considerable comfort from the knowledge that at least one counseling treatment can consistently make a whopping big difference on one particular kind of clinical problem, however narrowly defined.

In addition to obscuring the effects of a potentially powerful treatment, failure to respond to the Appropriate Treatment Myth can erode our under-

standing of why a particular effect did indeed emerge. To illustrate this point, let us consider the treatment of phobias. When desensitized subjects move a foot closer to test animals than alternatively treated controls, researchers commonly conclude that their theoretically relevant treatment has caused a reduction in fear. It is possible, however, that actual phobic characteristics of the subject pool may remain unchanged; the gain in fact may be due to certain contextual variables of the treatment which inadvertently altered skepticism levels or mistaken beliefs. For example, certain scenes in the desensitization hierarchy might underscore the notion that the animal is absolutely passive and harmless, or the scenes might contain new information such as the snake skin being cool and dry as opposed to wet and slimy. In this instance the researcher erroneously credits the theoretical framework of desensitization for producing a significant but trivial effect which might have been enormously magnified had an alternative treatment specifically addressed adaptive skepticism and/or mistaken belief.

The previous two consequences of failure to recognize the Appropriate Treatment Myth chronically occur. Linda Craighead has suggested to me the possibility of a third consequence. Perhaps the situation exemplified by "three studies reporting superiority of treatment A over B vis a vis four studies claiming victory for B over A," is really a function of ideosyncratic subject pool characteristics. In other words, the magnitude and direction of effect varies with the relevance of the treatment to the particular majority of the subjects. As the constellation of actual clinical problems changes from study to study, so might the outcome.

The Treatment Deployment Myth

The Treatment Deployment Myth is really a generic name for a number of interrelated delusions about how our counseling and psychotherapy treatments are implemented in the context of an experimental study. We are vastly mistaken if we think that our treatments are standardized, that they necessarily correspond to the theoretical principles on which they are supposed to be based, and that they are in fact received by the subjects in a given study. Let me briefly address each of these delusions.

1. The Standardized Treatment Delusion. Our literature suggests that we have few if any standardized treatments. Unlike the pharmacologist whose independent variables are capable of being held constant across time, geography, and publication outlet, counseling and psychotherapy interventions routinely vary on all conceivable dimensions. Even the originators of our treatment strategies rarely replicate the identical procedure from study to study, so it should hardly come as a surprise to find their students and peers in the research community making further alterations.

To illustrate, consider the rapid-smoking treatment for cigarette addiction. Studies purporting to test this seemingly circumscribed procedure have in fact varied on a) the numbers and nicotine ratings of cigarettes consumed, b) the amount of time smoking per trial and the number of trials per session, c) the number and spacing of treatment sessions, d) treatment group size, e) the presence or absence of therapeutic relationship qualities, homework assignments, booster sessions and so forth (see Danaher, 1977). As one might expect, the outcomes of these endeavors have also been quite variable. Similar diversity of course exists in the literatures of desensitization and modeling. If I were to ask members of this audience to fully describe the procedure commonly known as "behavioral rehearsal," I'm sure

dozens of differing operational definitions would emerge.

The effects of the Standardized Treatment Delusion are not entirely disadvantageous. For example, one might argue rather convincingly on the need to avoid prematurely freezing our treatment programs. In so doing we might shut off the opportunity for conceptual and pragmatic improvements, not to mention the possibility of serendipitous findings. On the other hand, capturing the concensus of our literature on the efficacy of a fluidly-defined technique is a bit like trying to pick up mercury with one's fingers. It's hard to get a hold of something to say.

The irony here is that our current methodological sophistication allows us the opportunity to enjoy the best of both worlds, consistency and diversity. Component, parametric, constructive, and dismantling analyses, for example, permit the replication of important treatment effects while at the same time allowing the investigator the opportunity to explore whatever other variables are of interest. Regrettably these roads remain relatively untraveled.

2. The Theory-Practice Congruence Delusion. Several philosophers of science have fully discussed the logical error of believing that the emergence of a particular hypothesized treatment effect confirms the underlying theory (e.g., Cook & Campbell, 1979; Mahoney, 1976; Popper, 1959; Weimer, 1976). There is a more fundamental delusion, however, that undergirds our literature, namely, the belief that our counseling interventions necessarily correspond to the theoretical principles on which they are supposed to be based. Let me cite several glaring examples of theory-practice incongruity.

We are all aware of the tenants of classical client centered therapy (Rogers, 1959; 1961). Unconditional positive regard, for example, by definition precludes the faintest hint of therapist-imposed values. Many of us are also familiar with Truax's (1966) illuminating analysis of

Carl Rogers in practice; Truax conclusively showed that Rogers differentially reinforced--via verbal conditioning--those kinds of client statements seen by Rogers as desirable. What then is client centered counseling? Is it what Rogers says he does (i.e., his theory)? Or is it what he in fact does (i.e., his practice)? From an empirical standpoint we can clean up the situation by either revising the theory of client centered therapy or by excluding all data produced by erratically behaving counselors including Rogers himself. Our literature suggests we've done neither.

In the foregoing example, we find the proponent of a technique in violation of his theoretical principles. Can we thus seriously expect antagonistic individuals to provide adequate representation of a given theory or practice in the context of their experiments? We all know of behaviorists who arrogantly label their placebo treatments as "client centered therapy" on the basis of superficial similarities while ignoring fundamental differences. Perhaps less well known or acknowledged is the large number of so-called "behavioral" projects conducted by individuals who seemingly haven't the foggiest understanding of the principles and practices they purport to examine. Walt Disney's skunk named "Flower" was still a skunk. Simply because a study claims to examine a given intervention does not mean that the intervention was in fact adequately examined.

My final example of theory-practice incongruity concerns those theoretical principles which seem to defy implementation in counseling practice even by the most well-versed and dispassionate of experimenters. The theory underlying negative reinforcement, for example, demands that the escape response (e.g., an adaptive target behavior) produce a cessation of the noxious stimulus. Yet in the counseling strategy labeled "covert negative reinforcement" the noxious stimulus (an unpleasant image) is terminated <u>before</u> the

adaptive behavior is begun. Similar implementation difficulties exist with other interventions such as coverant control, time out, and response cost (see Horan, 1979; Mahoney, 1974).

I wish there were a simple cognitive restructuring remedy for the Theory-Practice Congruence Delusion which pervades our professional literature. There is not. I take little comfort in the atheoretical cop-out offered by others: "Forget the theory," they say, "let the operations and emergent data speak for themselves." True enough in the short run, but eventually we must present to our consumer audience and to our contemporaries in other professions a set of coherent (albeit evolving) theoretical principles supported by data gathered in practice. It seems to me the time has come for counseling and psychotherapy editorial reviewers to pay less attention to issues such as the comparative merits of ANCOVA vs Repeated Measures ANOVA in a particular manuscript and focus more on the oftentimes missing link between the conceptual basis of a study and its implementation.

3. The Subject Receptivity Delusion. The first two delusions supporting the Treatment Deployment Myth concern matters which are to some degree under the control of the experimenter. The author of a study decides which version of a "standard" treatment he or she wishes to evaluate, and moreover determines whether or not the treatment corresponds to the principles on which it is supposed to be based. Authors do not necessarily control, however, what their subjects do with the treatment. In pharmacological research this problem is called "cheeking the pill " (instead of swallowing), and there are simple ways to deal with it. The field of counseling and psychotherapy, however, is not so fortunate.

To illustrate, much has been written about the stimulus control approach to the treatment of obesity. The logic of stimulus control rests on the

assumption that the eating behavior of obese subjects is essentially "out of control;" that is, they purportedly take large bites, eat rapidly, and let extraneous factors such as time of day and the availability of food determine how much is eaten. Apart from the fact that these propositions have come under some empirical assault (e.g., Mahoney, 1975), we have remarkably little evidence to support our further assumption that obese individuals who are given stimulus control training actually alter their eating style upon leaving the counseling cubicle. The stimulus control treatment of obesity typifies the perplexing situation in which a powerful treatment effect can be expected to occur in spite of the fact that subjects many routinely "cheek the pill."

The converse of this situation exists, of course, when a potentially powerful treatment is for some undetermined reason ignored by the subjects and a null effect ensues. In a recent component analysis of stress inoculation, for example, we found that self-instructions training was conspicuously ineffective on all outcome measures (Hackett & Horan, 1980). In contrast, two other categories of coping-skill training definitely proved their worth. A check on the independent variable manipulation, however, revealed that only half of the subjects who received self-instructions training actually put that training into practice. For the other two coping skill categories adherence to the treatment was nearly universal.

Independent variable manipulation analyses are routinely conducted in certain areas of education and psychology, but they are surprisingly rare in the counseling and psychotherapy literature. One would think that experimenters themselves might wonder if high percentages of subjects in the various treatment conditions were in fact doing what they were supposed to be doing (and not doing what they shouldn't be doing). Certainly, this sort

of information would greatly enhance our understanding of both null and positive effects.

Failure to rectify the three delusions supporting the Treatment Deployment Myth exacerbates the consequences of The Appropriate Treatment Myth; namely, we increase the risk of washing out treatment effects that might otherwise occur and we thoroughly obfuscate the meaning of those that do emerge. The final myth that I wish to address today, however, is perhaps the most problematic of all.

The Control Group Myth

In the counseling and psychotherapy literature authors invariably write as if their control groups had received "everything but" the experimental treatment. In point of fact "anything but" would be a more apt descriptor. This distinction is extremely important because the nature of the control condition has profound implications for the proper interpretation of what might appear to be a treatment effect. By The Control Group Myth I mean the common but erroneous belief that the inclusion of a randomly-assigned control condition allows one to determine whether or not the experimental treatment made a difference. Possibly so, but usually not. To place this issue in perspective a brief survey of counseling and psychotherapy control groups might be helpful.

Control group variations are legion. We have no-treatment controls and delayed treatment controls, each of which exist under varying levels of therapist contact, attention, concern, and hope for the future. We also have what are called "placebo controls." Placebo controls are supposed to be theoretically inert alternative treatments; however, when placebos are found to "work" (as is frequently the case) we rename them and build our careers on subsequent theory development.

Then there are minimal treatment controls, which involve the deployment

of active counseling interventions in quantities judged too small to make a difference, alternative treatment controls in which no amount of treatment is expected to make much of a difference, and standard treatment controls which pit our experimental interventions against the best, or at least modal, practices in the field. And the list goes on. We have counter-demand phases which allow the measurement of improvement in spite of posited subject expectations to the contrary, and what might be called "counter treatment controls" which theoretically produce deterioration in spite of posited subject expectations for improvement.

In the midst of all these variations and permutations it is easy to lose sight of why we bother with control groups in the first place. Investigators who use no treatment controls or delayed treatment controls are essentially asking "Did anything happen at all?" They view placebo influences as either nonexistant or trivial, or at least not important to distinguish from the effects of treatment per se. The problem here, of course, is that the treatment itself may be nothing more than a placebo.

In contrast, investigators who employ alternative activity control treatments would like us to believe that they have controlled the placebo problem. Aye, but here's the rub: The placebo phenomenon is not necessarily a function of what we in fact do to our subjects, but rather what they believe we are doing to them. Thus researchers who compare hum-drum bibliotherapy with fancy experimental treatments involving lab coats, lights, whistles and buzzers, routinely fail to realize that the emergent significant differences on outcome measures might well be a function of differential subject expectations for improvement. Equalizing "minutes-of-therapist-contact-time" by adding "verbal filler" does not resolve the problem and may even compound it. Such "psychobabble" could conceivably alienate subjects and erode whatever placebo influences the control treatment would otherwise muster!

A basic question that goes unanswered in most experimental studies of counseling and psychotherapy is essentially this: Did the subjects in the experimental and control conditions expect equivalent amounts of benefit? Kazdin and Wilcoxon's (1976) timely review of the desensitization literature, for example, found only 5 out of 98 projects that provided such assurances. Incidentally, only one of these projects unequivocally supported the efficacy of desensitization, and desensitization is often considered to be the most empirically validated treatment strategy in the field of counseling and psychotherapy.

Recent breakthroughs in our understanding of the biology and psychology of pain dramatically illustrate the need for counseling and psychotherapy researchers to ensure that their control treatments generate equivalent expectations for improvement. We now know, for example, the mere belief that one is receiving a pain killing drug actually causes one's body to produce and secrete endorphin, a form of opium (e.g., Levine, Gordon, & Fields, 1978) The placebo phenomenon thus has biochemical reference points!

The problem of differential subject expectations for improvement is known by a variety of names in the methodological literature. Some authors speak of differential demand characteristics; others refer to differential credibility or believability; still others list "rival hypotheses" that exist in spite of random assignment (see, Cook & Campbell, 1979, Jacobson & Baucom, 1977; Kazdin, 1979; Lieberman & Dunlap, 1979; Loney & Milich, 1978; and O'Leary & Borkovec, 1978). Fine lines of distinction might be drawn between each of these concepts, but it's not important to do so now. Generally speaking, in the counseling and psychotherapy literature we don't need a new name for the placebo phenomenon, just a more widespread realization that any so-called control treatment which does not generate equivalent subject expectations for improvement does not in fact control for the placebo phenomenon. And unless, of

course, we contain this widespread and powerful artifact, we cannot speak pridefully of a treatment effect regardless of the altitude of the obtained significance level.

I would like to thank you all for enduring my rendition of a contemporary Grimm Mythology. My points were essentially these: We have not proven that aggregated counseling and psychotherapy schools are effective at all, much less have we shown that they are equally effective. More fundamentally, however, our experimental subjects often do not receive treatments appropriate to their clinical problems, our treatments are frequently not deployed as purported, and finally our so called control groups rarely address one of the most powerful artifacts of all. In spite of these faulty beliefs and customs, we now have the methodological sophistication to lay a firm conceptual and empirical basis for our field. But unless we choose to purge these myths from our midst, the <u>practice</u> of counseling and psychotherapy will remain just that.

Footnotes

1. In so doing we are in effect saying that it doesn't matter that the subjects may be performing better because of the anxiety, the fact that they don't like the anxiety is reason enough to try and reduce it. This concession can cause a curious logical contradiction in those studies using GPA as an ancillary dependent measure!

References

Badia, P., Haber, A., & Runyon, R. P. Research problems in psychology. Reading, MA: Addison-Wesley, 1970.

Bandura, A. Effecting change through participant modeling. In J. D. Krumboltz & C. E. Thoresen (Eds.) Counseling Methods. N.Y.: Holt, Rinehart, & Winston, 1976. Pp. 248-265.

Bandura, A., Blanchard, E. B., & Ritter, B. The relative efficacy of desensitization and modeling approaches for inducing behavioral, affective, and attitudinal changes. Journal of Personality and Social Psychology, 1969, 13, 173-199.

Barber, T. X. Pitfalls in human research: Ten pivotal points. Elmsford, N.Y.: Pergamon, 1976.

Barrios, M. S. Repeating the mistakes of the past: A note on subject recruitment and selection procedures in analogue research on small animal phobias, AABT Newsletter, November 1977, 4(6), 19.

Bernstein, D. A. & Paul, G. L. Some comments on therapy analogue research with small animal "phobias." Journal of Behavior Therapy and Experimental Psychiatry, 1971, 2, 225-237.

Campbell, D. T., & Stanley, J. C. Experimental and Quasi-experimental designs for research. Chicago: Rand McNally, 1966.

Cook, T. D., & Campbell, D. T. Quasi experimentation: Design and analysis issues for field settings. Chicago: Rand McNally, 1979.

Cooper, A., Furst, J., & Bridger, W. A brief commentary on the usefullness of studying fear of snakes. Journal of Abnormal Psychology, 1969, 74, 413-414.

Cronbach, L. J., & Furby, L. How should we measure 'change' or should we? Psychological Bulletin, 1970, 74, 68-80.

Danaher, B. G. Research on rapid smoking: Interim summary and recommendations. Addictive Behaviors, 1977, 2, 151-166.

Ellis, A. Rational-emotive therapy: Research data that supports the clinical and personality hypotheses of RET and other modes of cognitive-behavior therapy. The Counseling Psychologist, 1977, 7, 2-42.

Eysenck, H. J. An exercise in mega-silliness. American Psychologist, 1978, 33, 515-517.

Fiedler, D., & Beach, L. R. On the decision to be assertive. Journal of Consulting and Clinical Psychology, 1978, 46, 537-546.

Gallo, P. S. Meta-analysis--A mixed meta-phor? American Psychologist, 1978, 33, 515-517.

Games, P. A. Limitations of analysis of covariance on intact group quasi-experimental designs. Journal of Experimental Education, 1976, 44, 51-54.

Glass, G. V. Primary, secondary, and meta-analysis of research. Educational Researcher, 1976, 5 (10), 3-8.

Glass, G. V. Reply to Mansfield and Busse. Educational Researcher, 1978, 7(1), 3.

Glass, G. V., & Smith, M. L. Reply to Eysenck. American Psychologist, 1978, 33, 517-519.

Hackett, G., & Horan, J. J. Stress inoculation for pain: What's really going on? Journal of Counseling Psychology, 1980, 27, 107-116.

Hackett, G., Horan, J. J., Buchanan, J. D., & Zumoff, P. Improving the exposure component and generalization potential of stress-inoculation for pain. Perceptual and Motor Skills, 1979, 48, 1132-1134.

Horan, J. J. Counseling for effective decision making: A cognitive-behavioral perspective. North Scituate, MA: Duxbury, 1979.

Horan, J. J., Hackett, G., Buchanan, J. D., Stone, C. I., & Demchick-Stone, D. Coping with pain: A component analysis of stress-inoculation. Cognitive Therapy and Research, 1977, 1, 211-221.

Jacobson, N., & Baucom, D. Design and assessment of nonspecific control groups in behavior modification research. Behavior Therapy, 1977, 8, 709-720.

Kazdin, A. E. Therapy outcome questions requiring control of credibility and treatment generated expectancies. Behavior Therapy, 1979, 10, 81-93.

Kazdin, A. E., & Wilcoxon, L. A. Systematic desinsitization and nonspecific treatment effects: A methodological evaluation. Psychological Bulletin, 1976, 83, 729-758.

Kazrin, A., Durac, J., & Agteros, T. Meta-meta analysis: A new method for evaluating therapy outcome. Behaviour Research & Therapy, 1979, 17, 397-399.

Kiesler, D. J. Some myths of psychotherapy research and the search for a paradigm. Psychological Bulletin, 1966, 65, 110-136.

Krumboltz, J. D. Future directions for counseling research. In J. M. Whitely (Ed.), Research in counseling: Evaluation and refocus. Columbus, Ohio: Merrill, 1968.

Levine, J. D., Gordon, N. C., & Fields, H. L. The mechanism of placebo analgesia. Lancet, September 23, 1978, 654-657.

Lieberman, L. R. & Dunlap, J. T. O'Leary and Borkovec's conceptualization of placebo: The placebo paradox. American Psychologist, 1979, 34, 553-554.

Loney, J., & Milich, R. Development and evaluation of a placebo for studies of operant behavioral intervention. Journal of Behavior Therapy and Experimental Psychiatry, 1978, 9, 327-333.

Lord, F. M. Statistical adjustments when comparing pre-existing groups. Psychological Bulletin, 1969, 72, 336-337.

Mahoney, M. J. Cognition and behavior modification. Cambridge, Mass.: Ballinger, 1974.

Mahoney, M. J. Fat fiction. Behavior Therapy, 1975, 6, 416-418.

Mahoney, M. J. Scientist as subject: The psychological imperative. Cambridge, MA: Ballinger, 1976.

Mansfield, R. S., & Busse, T. V. Meta-analysis of research: A rejoinder to Glass. Educational Researcher, 1977, 6(9), 3.

Meichenbaum, D. Cognitive-behavior modification. New York: Plenum Press, 1977.

Morganstern, K. P. Implosive therapy and flooding procedures: A critical review. Psychological Bulletin. 1973, 79, 318-334.

O'Leary, K. D., & Borkovec, T. D. Conceptual, methodological, and ethical problems of placebo groups in psychotherapy research. American Psychologist, 1978, 33, 821-830.

Popper, K. R. The logic of scientific discovery. N.Y.: Harper & Bros., 1959.

Presby, S. Overly broad categories obscure important differences between therapies. American Psychologist, 1978, 33, 514-515.

Rimland, B. Death knell for psychotherapy? American Psychologist, 1979, 34, 192.

Rogers, C. R. A theory of therapy, personality, and interpersonal relationships as developed in the client-centered framework. In S. Koch (Ed.), Psychology a study of science, Vol. III, Formations of the person and the social context. New York: McGraw-Hill, 1959. Pp. 184-258.

Rogers, C. R. On becoming a person. Boston: Houghton-Mifflin, 1961.

Rosenthal, R., & Rosnow, R. L. Artifact in behavioral research. N.Y.: Academic Press, 1969.

Scriven, M. Behavior therapies: What do they mean? Implications for theory, research, and practice. Panel Discussion--Symposium, Annual meeting of the American Educational Research Association, San Francisco, April 10, 1979.

Sherry, G. S., & Levine, B. A. An examination of procedural variables in flooding therapy. Behavior Therapy, 1980, 11, 148-155.

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. American Psychologist, 1977, 32, 752-760.

Stern, R., & Marks, I. M. Brief and prolonged flooding. Archives of General Psychiatry, 1973, 28, 270-276.

Truax, C. B. Reinforcement and non reinforcement in Rogerian psychotherapy. Journal of Abnormal Psychology, 1966, 71, 1-9.

Weimer, W. B. Psychology and the conceptual foundations of science. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Yerkes, R. M., & Dodson, J. D. The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology, 1908, 18, 459-482.

* Kazdin, A. E. & Wilson, G. T. Evaluation of behavior therapy: Issues, evidence, and research strategies. Cambridge, MA: Ballinger, 1978.